

(12) **UK Patent Application** (19) **GB** (11) **2 200 118** (13) **A**
(43) Application published 27 Jul 1988

(21) Application No 8701516

(22) Date of filing 23 Jan 1987

(71) Applicant
Allelix Inc.

(Incorporated in Canada)

**6850 Goreway Drive, Mississauga, Ontario,
Canada, L4V 1P1**

(72) Inventors

**Michael Alan Wosnick
Martin Sumner-Smith
Richard Warren Barnett
David Ivor Gwynne
Francis Paul Buxton
Nigel Armstrong Skipper
Roger Wayne Glasse-Davies**

(51) INT CL⁴
C12N 15/00

(52) Domestic classification (Edition J):
**C3H 649 653 654 734 B7E
U1S 1086 C3H**

(56) Documents cited
**GB A 2123005 GB A 2100737 GB A 2091271
EP A2 0196921 EP A2 0161866 EP A1 0134662
EP A1 0123928 EP A1 0121775 EP A1 0077109
WO A 83/04418 US 4666847**

(58) Field of search
**C3H
Selected US specifications from IPC sub-class
C12N**

(74) Agent and/or Address for Service
**Batchellor, Kirk & Eyles
2, Pear Tree Court, Farringdon Road,
London, EC1R 0DS**

(54) Synthetic chymosin-and prochymosin-encoding DNA segments

(57) Synthetic protein encoding DNA segments are disclosed which encode natural proteins but which have a base sequence whose codons are more preferred by an intended host cell than is the base sequence of the natural protein-encoding segment. Host transformation is conducted with vectors containing the synthetic coding segment. In a specific embodiment, yeast or filamentous fungi are transformed to express a synthetic segment which encodes prochymosin.

GB 2 200 118 A

SYNTHETIC CHYMOSIN- AND PROCHYMOSIN-ENCODING DNA SEGMENTS

The present invention relates generally to molecular biology. It pertains, more particularly, to synthetic genes and host transformation therewith.

BACKGROUND OF THE INVENTION

One goal of molecular biology research is to provide alternative, commercially viable sources of proteins and enzymes for industrial use. This goal is satisfied in part by manipulating cells which can be maintained economically to produce proteins otherwise too costly or impractical to produce. Recombinant DNA technology provides the tool by which the protein-encoding genes are transferred from one cell to another to obtain protein production in the desired host. While this has been accomplished on many occasions, there remains the goal of obtaining more efficient expression of the foreign gene in the transformed host.

In conventional rDNA techniques, a gene from one species is often used to transform a cell of a different species, genus, family or even kingdom, without regard for the compatibility of the gene with its foreign host or, more importantly, for the ability of the transformed host to express the foreign gene as efficiently as is possible. For example, synthesis of the protein encoded by the new gene may be hindered at the translational level if the required tRNA species are not present in sufficient quantity to meet the needs of translating a foreign gene efficiently.

In fact, it has been established that organisms do not utilize all possible codons with equal frequency. Given that almost all amino acids are encoded by two or more codons, e.g. serine is encoded by six different codons, threonine by four different codons etc., this selective codon usage i.e. codon

bias, can affect the level at which the foreign gene is expressed particularly when the source of the gene and the host cell are genetically unrelated or, at least, exhibit distinctly different codon bias.

This codon bias may be present, in a subtle way, in all genes of a given organism, but is most pronounced in genes which are expressed to very high levels. The codon selection may be extremely severe in these cases. For example, one study (Bennetzen and Hall, 1982) on the yeast highly expressed genes for alcohol dehydrogenase I (ADH-I) and glyceraldehyde-phosphate dehydrogenase showed that 96% of the amino acid residues were encoded by only 25 out of the possible 61 coding triplets. This "streamlining" of the codon usage makes it possible to express these genes to very high levels, since the time taken to decode the mRNA will clearly be the minimum possible, with no delays to accomodate the effects of relatively rare tRNAs.

The codon bias for highly expressed genes appears to be a distinct feature in all genomes. However, the codon bias from organism to organism is not necessarily the same. For example, although E. coli exhibits a very severe codon bias for highly expressed genes such as those encoding either the major lipoprotein or the elongation factors (Guoy and Gautier, 1982), the actual codons used may be different from those which are highly preferred in another organism, such as yeast (Bennetzen and Hall, 1982).

Although the art suggests that codon bias disparity exists between different organisms, this knowledge has yet to be applied in a practical way with a view to enhancing gene expression in transformed hosts.

DESCRIPTION OF THE PRIOR ART

In the art, attempts at enhancing levels of gene expression focus almost entirely on the identification,

manipulation and modification of gene elements such as promoter regions, secretion signals and termination regions or other gene elements which exert some transcription or translation-controlling function over the protein-encoding region of a gene which is usually genome-derived or cDNA based.

A typical example is the approach taken to develop rDNA-based biological systems for production of the enzyme known as chymosin (also known as rennin).

Chymosin is an aspartyl protease (EC 3.4.23.4) which is normally found in the fourth stomach of the unweaned calf, where it functions in the clotting of milk. It is a secreted protein, initially synthesized as a longer precursor preprochymosin. Preprochymosin has a 16 amino acid signal peptide at its amino terminus which is cleaved, upon secretion, to yield the zymogen prochymosin.

The clotting property of chymosin is very important commercially, since chymosin is the preferred milk coagulant for the cheese industry. Unfortunately, because of the decreasing world-wide demand for veal, the supply of calf stomach from which to recover the enzyme is declining. Accordingly, recombinant DNA technology has been applied to provide a sufficient, stable supply of chymosin.

Cloning of the chymosin gene in E. coli was reported by Beppu et al in J. Biochem 90: 901-904 (1981). The nucleotide sequence of calf prochymosin cDNA was reported later by Beppu in J. Biochem 91 1085-1088 (1982). The development of chymosin vectors progressed with the disclosure of an E. coli expression plasmid having the lacUV5 promoter (1982) or the tryptophan promoter.

Transformation vectors suitable for expression of prochymosin by yeast are described in Gene 24 pp 1-14. (1983)

which discloses a vector in which prochymosin cDNA is under the influence of the yeast PGK gene; in Science, Vol. 229 pp 1219-1223 (September 1985) which describes use of the yeast invertase secretion signal in a chymosin cDNA vector and in Gene 27 pp 35-36 (1984) which describes a vector in which chymosin cDNA is under control of the yeast GAL1 promoter and contains the yeast SUC2 transcriptional terminator fragment.

In all of the above disclosures, the chymosin or prochymosin fragment on the vector is derived from bovine cells i.e. is either genomic DNA, or is a cDNA copy of the corresponding messenger RNA. No experiments are described which attempt to reconcile the codon bias of the new host i.e. yeast or E. coli as opposed to bovine stomach cells, with the codon usage in the chymosin encoding fragment contained on the vector. Similar approaches have been taken in creating hosts engineered to express other foreign proteins. Although some gene elements are manipulated and substituted, the protein coding segment used is a natural, unaltered segment.

SUMMARY OF THE PRESENT INVENTION

In the present invention, use is made of protein coding regions which code for authentic proteins or polypeptides but which have been wholly or partially synthesized according to the codon bias of the cell which is to be transformed. Once synthesized, the coding region may be coupled with other gene segments, as desired e.g. a promoter region, a secretion signal sequence, a termination region etc. and combined if necessary with a suitable marker, replication origin and the like in order to form a suitable transformation vector. Since, by design, the synthetic coding region which encodes the protein of interest is comprised of codons for which the eventual host has a preference, translation of the synthetic protein coding region is able to progress at an enhanced rate compared with the rate at which translation of an authentic but foreign coding region would occur in the same host.

Thus, according to one aspect of the present invention, there is provided a protein-encoding DNA segment having a base sequence optimized for expression in a host to which said protein is foreign.

In accordance with further aspect of the present invention, there are provided vectors containing the protein-encoding DNA segment, and host cells transformed therewith.

In accordance with aspects of the present invention which are preferred herein, there is provided a chymosin- or prochymosin-encoding DNA segment having a base sequence optimized for expression in a foreign host. Vectors containing these synthetic coding regions and hosts transformed therewith are also within the scope of the present invention.

In the present invention, the protein coding region is optimized by substituting codons defined by the natural nucleotide sequence with codons which are preferred by the intended recipient i.e. host cell which is to be transformed. All or portions of the optimized coding region may be produced synthetically, the portions then being coupled to the remaining natural segment, if any, to form an entire optimized coding region. In this sense, the synthetic coding region and the authentic coding region are analogous in that they encode the same or substantially the same amino acid sequence. (By "substantially the same", it is meant that the amino acid sequences are functionally equivalent in terms of utility or activity.) However, some number of codons defined by the synthetic coding region have been altered by comparison with the natural coding region so as to reflect the codon bias of the intended host.

It is believed that supplanting of even one codon for a more preferred codon will affect the level at which the coding

region is expressed by enhancing that level to some extent. It may be more desirable to alter the codon usage to a greater extent, however with maximum advantage being realized when all codons of a given coding region are optimized for the host.

In order to synthesize a protein coding segment optimized for a particular host, knowledge of the codon bias of the intended host is required. Provided that the nucleotide sequence of at least a few genes whose protein products are highly expressed by the natural host are known, this information can be calculated. From the nucleotide sequence, it is currently a simple matter to determine the corresponding amino acid sequence and to tabulate the frequencies at which the codons are used. When codon usage is compared with amino acid usage, a pattern of codon bias emerges. With the benefit of this data, a coding region may be synthesized which utilizes only codons preferred by the particular host whose codon preference has been determined. Indeed, once this pattern is determined, the application of the principle is simplified. All that is then required is to identify the amino acid sequence of the protein encoded by the foreign gene with which the host is to be transformed. Once the amino acid sequence of the protein encoded by the foreign gene is determined, a gene optimized for expression by the new host can be synthesized by incorporating only those codons which are the codons preferred by the new host for the corresponding amino acids, where enhanced expression is desired.

Despite the specific tailoring of the coding region with respect to one particular host, it has been found that the optimized coding region can be expressed not only in the particular host species for which it was designed, but also in other hosts. For example, it has been found that a protein-coding region optimized specifically for expression in yeast can also be expressed in filamentous fungi without further alteration. Accordingly, the optimized protein-encoding region

DNA segments of the present invention are useful in obtaining expression from hosts other than those specific hosts for which the optimized segment was originally intended.

The hosts with which the present invention is concerned primarily are the fungi, including the filamentous fungi and the yeasts. Given the general principle underlying the present invention, it will be appreciated that other hosts are included within its scope.

Thus, the present invention provides a synthetic protein-encoding DNA segment which is analogous to a natural such segment wherein the synthetic segment is optimized for expression in a particular host cell foreign to the natural segment.

As "protein encoding DNA segments" there may be mentioned signal sequences, sequences encoding mature proteins, entire coding regions, segments thereof including polypeptides and the like.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Given their well established role in industrial application of recombinant DNA technology and efficient secretive capacity, the yeasts are most preferred as hosts for the optimized sequences including Saccharomyces sp. such as carlbergensis and cerevisiae and the species of the genera pichia and Candida. Filamentous fungi e.g. Aspergillus sp. are also candidates for application of the principle expounded herein, particularly A. niger and A. nidulans. Of particular preference is the use of Saccharomyces cerevisiae as host.

The codon bias of S. cerevisiae was determined using the procedure described generally above. The results are shown in Table I below:

TABLE I

Codon Bias of S. cerevisiae in Highly Expressed Genes

Amino Acid	Codon	% Occurrence Yeast
Gly	GGG	0
Gly	GGA	1
Gly	GGT	97
Gly	GGC	2
Glu	GAG	2
Glu	GAA	98
Asp	GAT	29
Asp	GAC	71
Val	GTG	0
Val	GTA	0
Val	GTT	52
Val	GTC	48
Ala	GCG	0
Ala	GCA	0
Ala	GCT	76
Ala	GCC	24
Arg	AGG	0
Arg	AGA	92
Arg	CGG	0
Arg	CGA	0
Arg	CGT	8
Arg	CGC	0
Lys	AAG	91
Lys	AAA	9
Asn	AAT	3
Asn	AAC	97
Met	ATG	100
Ile	ATA	0
Ile	ATT	43
Ile	ATC	57
Thr	ACG	1
Thr	ACA	0
Thr	ACT	41
Thr	ACC	59
Trp	TGG	100

TABLE I (cont'd)

Codon Bias of S. cerevisiae in Highly Expressed Genes

Amino Acid	Codon	% Occurrence Yeast
Cys	TGT	95
Cys	TGC	5
Tyr	TAT	1
Tyr	TAC	99
Phe	TTT	4
Phe	TTC	96
Ser	TCG	1
Ser	TCA	2
Ser	TCT	51
Ser	TCC	45
Ser	ACT	0
Ser	AGC	1
Gln	CAG	0
Gln	CAA	100
His	CAT	3
His	CAC	97
Leu	CTG	0
Leu	CTA	2
Leu	CTT	0
Leu	CTC	0
Leu	TTG	91
Leu	TTA	7
Pro	CCG	0
Pro	CCA	93
Pro	CCT	6
Pro	CCC	1

This codon preference data was derived from the nucleotide sequences of several very highly expressed yeast genes: phosphoglycerate kinase (Hitzeman et al., 1982) alcohol dehydrogenase (Bennetzen and Hall, 1982), enolase A and B (Holland and Holland, 1981), glyceraldehyde-3-phosphate dehydrogenase A and B (Holland and Holland, 1981) and pyruvate kinase (Burke et al., 1983).

Evidence that the codon bias exists in S. cerevisiae is clear in the Table. For example, although glycine has four synonymous codons, the codon GGT is preferred dramatically; while arginine has six synonymous codons, the AGA triplet is almost always preferred and of the six synonymous codons for leucine, TTG is by far the most often utilized. It is believed that this codon bias correlates directly with the availability of complementary tRNA species within the tRNA population of the S. cerevisiae genome. It will be appreciated therefore that when a foreign gene is inserted into the S. cerevisiae genome which does not possess the codons preferred by this host, there will be disparity at the translational level of protein synthesis, between the preferred codons and the availability of appropriate tRNA species. This can impact on the level at which the product of the foreign gene is expressed by the host.

A striking example of differences in codon preference is observed when the codon preference in highly expressed yeast genes is compared with the codon preference in the prochymosin gene. As described above, chymosin (rennin) is an aspartyl protease (EC 3.4.23.4) which is normally found in the fourth stomach of the unweaned calf. Thus, the yeast genome and the prochymosin source are essentially unrelated genetically. A comparison of codon usage in yeast genes and the actual codon usage in the prochymosin gene appears in Table II below:

TABLE II

Comparison of Codon Bias:
Bovine Prochymosin vs. Yeast Highly Expressed Genes

<u>Amino Acid</u>	<u>Codon</u>	<u>% Occurrence</u>	
		<u>Bovine Prochymosin</u>	<u>Yeast</u>
Gly	GGG	39	0
Gly	GGA	3	1
Gly	GGT	9	97
Gly	GGC	48	2
Glu	GAG	86	2
Glu	GAA	14	98
Asp	GAT	10	29
Asp	GAC	90	71
Val	GTG	55	0
Val	GTA	7	0
Val	GTT	10	52
Val	GTC	28	48
Ala	GCG	5	0
Ala	GCA	0	0
Ala	GCT	16	76
Ala	GCC	79	24
Arg	AGG	67	0
Arg	AGA	11	92
Arg	CGG	0	0
Arg	CGA	11	0
Arg	CGT	0	8
Arg	CGC	11	0
Lys	AAG	60	91
Lys	AAA	40	9
Asn	AAT	25	3
Asn	AAC	75	97
Met	ATG	100	100
Ile	ATA	5	0
Ile	ATT	9	43
Ile	ATC	86	57
Thr	ACG	8	1
Thr	ACA	21	0
Thr	ACT	17	41
Thr	ACC	54	59

TABLE II (cont'd)

Comparison of Codon Bias:
Bovine Prochymosin vs. Yeast Highly Expressed Genes

<u>Amino Acid</u>	<u>Codon</u>	<u>% Occurrence</u>	
		<u>Bovine Prochymosin</u>	<u>Yeast</u>
Trp	TGG	100	100
Cys	TGT	57	95
Cys	TGC	43	5
Tyr	TAT	23	1
Tyr	TAC	77	99
Phe	TTT	30	4
Phe	TTC	70	96
Ser	TCG	11	1
Ser	TCA	3	2
Ser	TCT	11	51
Ser	TCC	28	45
Ser	ACT	11	0
Ser	AGC	36	1
Gln	CAG	88	0
Gln	CAA	12	100
His	CAT	33	3
His	CAC	67	97
Leu	CTG	64	0
Leu	CTA	6	2
Leu	CTT	6	0
Leu	CTC	21	0
Leu	TTG	3	91
Leu	TTA	0	7
Pro	CCG	19	0
Pro	CCA	6	93
Pro	CCT	6	6
Pro	CCC	69	1

The prochymosin data was derived from the sequence of bovine prochymosin B as published by Harris et al., 1982.

It is clear from Table II that while yeast have a bias for the AGA codon, the prochymosin gene utilizes the AGG codon to code for the same amino acid, glycine. The preferred

glutamine residue in the prochymosin gene is CAG whereas in yeast genes, it is CAA. Other differences in codon preference are equally apparent from Table II.

It will be appreciated from the above that the coding region of the natural prochymosin gene is not well suited for expression in yeast, in terms of codon usage.

To prepare an analogous gene, in which codon usage is optimized for expression in yeast, nucleotide synthesizers are preferably employed. Entire genes may be constructed if desired, although, since only some of the codons need be substituted to enhance expression to some extent, synthetic segments in which optimized codons are utilized may be prepared and used to replace the corresponding native portion of the selected gene. The present inventors have been able to synthesize an entire gene consisting of over 2,200 nucleotides suggesting that, with modern techniques, abilities to synthesize long genes entirely should not be limiting to the application of the principle of the invention.

It should be recognized that, when synthesizing the optimized coding region it is possible to delete codons or to substitute one or more codons of a coding region by other than the most preferred codons to produce a structurally modified polypeptide but one which has substantially the same activity or utility.

The synthetic coding region particularly preferred herein is a prochymosin-encoding segment which is described in greater detail in the examples. Since, in constructing the synthetic prochymosin coding region, a synthetic chymosin coding region is also created, it will be appreciated that the present invention also comprises a chymosin-encoding DNA segment having a base sequence which is optimized for expression in a foreign host. In constructing this coding region, codons preferred by

S. cerevisiae were incorporated for the most part although other, less preferred codons were employed sparingly in order to incorporate restriction sites within the coding region at locations which permit convenient ligation of other gene segments to form a suitable transformation vector containing the optimized coding region.

Transformation vectors suitable herein will incorporate a promoter region coupled with the optimized coding region so as to regulate transcription of the coding region. Since the preferred hosts are yeast and the filamentous fungi, the promoter will preferably be any one of the promoters whose utility in those hosts has been established. Where the intended host is a yeast, e.g. S. cerevisiae or S. carlsbergensis, the promoter may be selected from the promoter regions of the phosphoglucokinase gene, or the GAL1, GAL7, GAL10, invertase or melibiase genes. The promoter region of the melibiase gene described in International publication number WO86/03777 published July 3, 1986 is preferred for use in yeast transformation vectors.

Where the intended host is a filamentous fungus, such as Aspergillus sp., including the niger and nidulans species which are preferred filamentous fungus hosts, the promoter region is preferably derived from any one of the following Aspergillus genes; glucoamylase, alcohol dehydrogenase I and aldehyde dehydrogenase. These promoter regions, and vectors containing them are described in greater detail in co-pending U.S. patent application serial number 811,404 filed December 20, 1985 which is incorporated herein by reference.

The transformation vectors suitable herein preferably comprise a secretion signal sequence in reading frame with and preferably directly fused with the optimized coding region, serving to signal secretion of the protein once translated. Suitable signal sequences for yeast hosts include the secretion

signal of the melibiase gene described in the international publication cited above. Signal sequences useful in filamentous fungi include the signal sequence of the glucoamylase gene or a synthetic consensus signal sequence, both of which are described in the U.S. patent application cited above.

It will be appreciated, as well, that the natural signal associated with pre-prochymosin i.e. the "pre-" region may be used as a secretion signal in the vectors of the present invention. More preferably, however, the "pre-" region base sequence is optimized in the manner set forth herein with respect to the prochymosin and chymosin base sequences.

In a particularly preferred embodiment, the selected optimized coding region encodes prochymosin although it is emphasized that the optimized coding region of chymosin i.e. the optimized prochymosin region from which the first 42 5' codons have been removed (see Figure 1, arrow A) may also be used to obtain expression of mature, active chymosin. The prochymosin coding region is incorporated, preferably, on any one of a number of suitable vectors available from the American Type Culture Collection and used to transform the selected host. For example, the pGL2 plasmids comprise the promoter region and signal sequence of the A. niger glucoamylase gene. Accordingly, the optimized prochymosin coding region can be spliced into the vector, in reading frame with the signal sequence and used to transform a filamentous fungus host, preferably either A. nidulans or A. niger. Aspergillus sp. transformation can also be accomplished using the plasmid pALCAL1S which comprises the promoter of the alcohol dehydrogenase I gene of A. nidulans and a synthetic consensus signal sequence, once the optimized prochymosin gene is appropriately incorporated therein.

Yeast transformation is most preferably accomplished using the plasmid p4 which comprises the promoter region and signal sequence of the melibiase gene of S. cerevisiae, once the

optimized chymosin region is appropriately incorporated into the p4 vector.

The specific plasmids described above i.e. the pGL2 series of plasmids, pALCA1S and p4 are on deposit with the American Type Culture Collection in Rockville, Maryland, U.S.A. as follows:

<u>Plasmid</u>	<u>ATCC Accession #</u>	<u>Deposit Date</u>
pGL2A	53365	December 16, 1985
pGL2B	53366	December 16, 1985
pGL2C	53367	December 16, 1985
pALCA1S	53368	December 16, 1985
p4	53360	December 16, 1985

Plasmids containing the optimized prochymosin coding region, the construction of which is described in detail hereinafter, were deposited with ATCC in January, 1987 in E. coli host DH1 and have been allotted the following accession numbers:

PMV-1/CHYM 105	ATCC 67294
pALCA1S/CHYM 103	ATCC 67295
pGL2C.CHYM 101	ATCC 67296

Yeast hosts have been successfully transformed with plasmid PMV-1/CHYM 105 and filamentous fungus hosts have been successfully transformed with plasmids pGL2C/CHYM 101 and pALCA1S/CHYM 103.

An embodiment of the invention is described hereinafter by way of example only with reference to the accompanying drawings in which:

Figure 1 represents, on three sheets, the nucleotide sequences of both the authentic and the optimized prochymosin A coding region (lower nucleotide sequence) as well as the amino acid sequence corresponding thereto, using conventional abbreviations for nucleotides and amino acids;

Figure 2 illustrates, on 5 sheets, the plurality of synthetic oligonucleotides which were coupled to form the optimized prochymosin coding region;

Figure 3 illustrates schematically the creation of plasmid pCHYM1A(Y)B2 and pCHYM1AB2-3 which comprise portions of the optimized prochymosin coding region;

Figure 4 illustrates the creation of plasmid pCHYM345-21 which comprises another portion of the optimized prochymosin coding region;

Figure 5 illustrates, in general terms, the creation of transformation vectors which incorporate the entire optimized prochymosin coding region;

Figure 6 is a plasmid map of pMV-1/CHYM105;

Figure 7 is a plasmid map of pGL2C/CHYM101; and

Figure 8 is a plasmid map of pAlcA1S/CHYM103.

In the example which follows, the bovine prochymosin gene and, consequently, the chymosin segment thereof, has been totally chemically synthesized in order to incorporate a codon bias which matches that of the yeast, Saccharomyces cerevisiae. This will unblock a potentially rate-limiting translation step and permit a level of expression that might not otherwise be attainable using a cDNA-based copy of the gene.

The difference in codon usage between the bovine prochymosin gene and a group of highly expressed yeast genes, is indeed quite dramatic. Referring to Table II, supra, it is readily seen that many of the codons in the natural prochymosin gene are among the worst choices for efficient expression in yeast, especially those for Gly, Glu, Ala, Arg, Gln, Leu and Pro, and to a marked but lesser extent for such amino acids as Val, Lys, Ile, Cys and Ser.

By chemically synthesizing the entire coding region of the gene for bovine prochymosin, all of those codons which would otherwise result in less efficient prochymosin production in yeast are replaced. This represents a very significant improvement to the conventional strategy of expressing a cDNA or genomic clone version of the particular gene, especially in a micro-organism which is capable of high level expression when not otherwise constrained by poor codon selection.

The codon-optimized synthetic bovine prochymosin gene was originally derived from the published amino acid sequences of prochymosin B (Pederson et al., 1975; Foltmann et al., 1979). This full peptide sequence was "reverse translated", by a standard computer program (Devereux et al., 1984) back into very limited codon assignments. For each amino acid, only the most frequently used yeast codon, from yeast highly expressed genes, (i.e. the codons with the highest usage frequencies in Table I, supra) was utilized. This produced a full sequence encoding bovine prochymosin B which was totally optimized for yeast high level expression.

This gene sequence is but one example of the invention herein claimed. For the particular embodiment described herein, however, a few further (minor) modifications were made.

Some of the nucleotide positions were further changed to allow the creation of desirable restriction enzyme

recognition sites, which were deemed useful for ease of cloning, sequencing and other subsequent manipulations. In addition, a few positions were changed for the reverse reason, namely to remove restriction enzyme sites that might prove troublesome for these manipulations. It must be stressed however, that these few additional modifications were made without in any way compromising the underlying principle of using high bias codons. In those cases where restriction sites were added or deleted, the substituted codon would therefore no longer be the one with highest frequency, but would be replaced by one with slightly lower preference. In no case was a codon which is decidedly unfavourable ever used.

Additionally, in the particular example described herein, the segment encoding prochymosin B was further modified to become a segment which encodes prochymosin A (Moir et al., 1982). These isozymes differ by only three amino acid substitutions i.e. of the 366 amino acids which define chymosin, Val¹³⁹, Ser²¹⁶ and Gly²⁸⁶ in prochymosin B correspond to Leu¹³⁹, Cys²¹⁶ and Asp²⁸⁶, respectively, in prochymosin A. It is to be noted however, that since there is nothing fundamentally different between prochymosin A and B, the coding region which encodes the B isozyme could just as well have been used.

The coding region of bovine prochymosin A is shown in Figure 1. Each row represented in Figure 1 identifies the amino acid sequence (top line), the nucleotide sequence of the natural coding region (middle line) and the nucleotide sequence of the optimized coding region (bottom line). In addition, restriction sites are identified either by " - " to indicate a site which has been deleted in the optimized segment or by " + " to indicate where a particular site has been added. It will be readily seen from Figure 1 that the optimized gene is very significantly different in sequence from the natural counterpart. There are in fact 264 substitutions out of a total

of 1095 coding nucleotides, or greater than 24% difference. A significant number of these differences are in the third positions of many codons, as expected. However, many wholesale codon changes have also been made, in the cases of Serine, Arginine and Leucine which are each encoded by two different families of codons which differ in more than the third position.

Figure 1 also indicates the coding region of bovine chymosin A since this region is contained within the prochymosin region illustrated. The synthetic chymosin coding region is represented by the base sequence beginning at arrow A on Figure 1 and terminating where indicated for prochymosin vectors containing it, represent an additional, preferred embodiment of the present invention.

The optimized gene for prochymosin A, as shown in Figure 1 was therefore synthesized using several modifications to the accepted automated procedures for synthesizing oligodeoxynucleotides, resulting in the capability to make longer oligonucleotide chains. The overall design strategy for the assembly of the synthetic gene made use of fewer oligonucleotides of longer average length than is currently the norm. This enabled construction of the entire gene in fewer overall steps.

The entire gene was comprised of 28 separately synthesized oligonucleotides (14 pairs, whose sequences were precisely complementary, except for terminal overhangs). These oligonucleotides ranged in length from 59 to 102 nucleotides. The full sequences and designations of these oligonucleotides are shown in Figure 2 referenced in more detail hereinafter. The oligonucleotides were separated into subassemblies 1 through 5. The oligonucleotides comprising each individual subassembly were assembled separately into an appropriate (commercially available) M13mp vector. In some cases (subassemblies 1, 2 and 3), small extensions were added to the complementary part of

oligonucleotides comprising one end of the final subassembly, strictly to facilitate the intermediate cloning and sequencing steps. These small extensions were left behind when the actual prochymosin gene portion was subsequently removed.

The prochymosin encoding portions of the various subassemblies, were then excised out of their intermediate plasmid hosts, and assembled into the final gene. The prochymosin gene was placed under the control of one of three different promoters and secretion signals: the melibiase promoter and signal for expression and secretion in the yeast, Saccharomyces cerevisiae, i.e. plasmid p4 ATCC 53360 and either the Aspergillus niger glucoamylase promoter and secretion signal i.e. plasmid pGL2C, ATCC 53367 or the Aspergillus nidulans alcohol dehydrogenase promoter and a synthetic signal for expression and secretion by Aspergillus i.e. plasmid pALCAL5, ATCC 53368. The overall strategy for the assembly of the gene is shown in Figures 3, 4, and 5.

Synthesis of Oligonucleotides

a) Materials

Oligonucleotides were prepared on a Biosearch SAM ONE Series II or Applied Biosystems 380B DNA synthesizer. Methyl-N,N-diisopropylaminophosphoramidites were obtained from Applied Biosystems (ABI) or Beckman Instruments. B-Cyanoethyl-N,N-diisopropylaminophosphoramidites were from ABI, American Bionetics (ABN), or Biosearch Inc. Nucleoside-derivatized controlled pore glass supports (CPG) were purchased from ABI or ABN. "Low-loaded CPG" was purchased from ChemGenes Corp. Bis-(B-cyanoethyl)-N,N-diisopropylaminophosphoramidite was obtained from ChemGenes. All other synthesis grade solvents and reagents were purchased from ABI or were prepared in-house according to ABI protocols. Tetrazole was obtained from either Cruachem or Aldrich Chemicals. Acetic anhydride (ACS reagent

grade), dichloroacetic acid (DCA, 99+%), 4-dimethylaminopyridine (DMAP, 99%), 2,6-lutidine (97%) and iodine (Gold Label) were purchased from Aldrich. Tetrahydrofuran (THF) and acetonitrile (MeCN, both HPLC grade) were purchased from Caledon Laboratories. The acetonitrile was rigorously dried by refluxing over calcium hydride and was distilled fresh just prior to use.

b) Synthesis

All of the prochymosin gene oligonucleotide components were synthesized using phosphoramidite chemistry (Beaucage and Carruthers, 1981; McBride and Carruthers, 1983) on either the Biosearch or ABI instrument according to modified Biosearch or ABI protocols, respectively. These modifications are described below.

Syntheses carried out on the Biosearch SAM ONE were run on a 0.3 μ mole scale of starting nucleoside instead of the standard 1 μ mole scale. The standard 20 minute AMIDITE program was used. This program utilized a 2.5 minute coupling with 25-30 μ moles of phosphoramidite (83- to 100-fold molar excess). Dichloroacetic acid (DCA, 2.5% in dichloromethane) was used instead of the stronger trichloroacetic acid (TCA) for removing the dimethoxytrityl protecting groups (detritylation step). The CPG-linked oligonucleotides were cleaved manually, directly in their columns by ammonium hydroxide, at room temperature, then completely deprotected by heating at 55°C for 12-16 hours.

Syntheses on the ABI 380B were carried out using a new ABI small-scale synthesis program (ssb003). The synthetic scale was further reduced from 0.2 μ moles to 0.1-0.15 μ moles of starting nucleoside thereby increasing the effective molar excess of phosphoramidite from 25-fold, to 33- to 50-fold (for 5 μ moles amidite per coupling). For syntheses of very long oligonucleotides (>60mers) the ChemGenes "low-loaded CPG" (5-10

μmoles nucleoside/gram CPG) was used instead of the usual 25-35 μmoles/gram CPG. The DCA treatment was increased from 5 x 10 sec. to 7 x 10 sec to ensure complete reaction during long oligonucleotide syntheses. Coupling times were increased from 35 sec. to 60-90 sec. to increase coupling efficiency. Oligonucleotides were cleaved automatically from the CPG on the synthesizer, then deprotected as above.

Two minor modifications were made to the 380B synthesizer to expand its synthetic capacity. A fourth acetonitrile reservoir was added to maximize the wash capacity. The scb003 synthesis program was altered to access each of these acetonitrile reservoirs equally during a cycle. In order to increase the solvent/reagent waste collection capacity the standard 2 litre waste bottle was modified to drain continuously into a 20 litre reservoir.

c) Purification

The oligonucleotides were purified by standard polyacrylamide gel electrophoresis methods (ABI User Bulletin No. 13, 1984), using 8-10% gels (1.5mm x 40cm x 16cm). Half of the crude oligomer was loaded into 2-4 10mm wells and electrophoresed for 6-8h at 400-500 volts. Product bands were visualized by UV-shadowing, excised from the gel, and the oligonucleotide electroeluted from the gel slices using an electroelution apparatus (International Biotechnologies Inc.). The oligonucleotides were eluted for 15-30 minutes at 120 volts into a 10M ammonium acetate solution, then desalted on C18 Sep-Pak cartridges (Waters Associates). Yields were determined by the absorbance at 260nm (OD260 units) on a Beckman DU8B spectrophotometer.

Enzymatic 5'-Phosphorylation of Oligonucleotides

a) Analytical Scale Phosphorylation with [^{32}P]-ATP

Small scale phosphorylations were used to check the purity of the oligonucleotides, and to establish estimates of the amounts of OD260 material that was actually capable of being phosphorylated. This latter determination was used to adjust the molar concentrations of oligonucleotides in the subsequent ligations.

Reactions were performed in buffer containing 50mM Tris-HCl (pH 7.5), 10mM MgCl_2 , 10 mM DTT, and 1.0mM Spermidine. Mixtures contained 1 pmol/ μl of oligonucleotide, 0.2 pmol/ μl [γ - ^{32}P] ATP (3uCi/pmol), 25 pmol/ μl of ATP, and 5-10 units of T4 polynucleotide kinase (P.L. Biochemicals). Incubation was for 40 min. at 37°C, followed by 10 min. at 65°C.

The labelled oligonucleotides were checked for purity by electrophoresis, under denaturing conditions, through a gel of 10% polyacrylamide containing 7M urea. The effective oligonucleotide concentration was determined by relating the actual number of Cerenkov counts recovered per band, to the theoretical amount expected, based upon the amount of O.D. 260 material, and assuming 100% efficiency for 5'-end labelling.

b) Large Scale Phosphorylation with [^{32}P]-ATP

For subassembly 2, a larger scale phosphorylation was performed. This was identical to that described above, except that the oligonucleotide concentration was quadrupled to 4 pmol/ μl , and the ATP concentration was increased to 100 pmol/ μl .

Vector Preparation

Each of the various subassemblies described below were cloned into one of either M13mpl0, M13mpl1, M13mpl9, or pBR322

(all of which are commercially available). These vectors were digested with the appropriate restriction enzymes, as described below, all according to manufacturers' specifications. Dephosphorylation of 5'-termini, when utilized, was accomplished using commercial preparations of either calf intestinal alkaline phosphatase or bacterial alkaline phosphatase, according to manufacturers' specifications. Vector fragments to be ligated to oligonucleotide subassemblies were purified by electrophoresis through gels of low-melting agarose (Bio-Rad), followed by isolation of the DNA by phenol extraction of the melted gel slice, according to widely accepted procedures (e.g. Maniatis et al., 1982).

Ligation of Oligonucleotide Subassemblies

a) "Shotgun Ligation"

In a typical experiment, phosphorylated oligonucleotides comprising one entire subassembly, were mixed together in annealing buffer (50mM Tris-HCl (pH 7.5), 10mM MgCl₂), at a final concentration of approximately 100 nM each (i.e. 1-5 pmol of each oligonucleotide in a total volume of 10-50 µl). The mixture was heated to 95°C for 5 min., followed by slow cooling to room temperature. The appropriate vector was then added, in proportions ranging from about 1/10 molar to approximately equimolar, with respect to the oligonucleotides added originally (i.e. for 2 pmol each of oligonucleotides, from 0.2 pmol to 2.0 pmol of vector was added). To this mixture was then added 1/10 vol of 10x ligation buffer (10x = 500mM Tris-Cl, pH 7.5, 100mM MgCl₂, 200mM DTT, 10mM ATP, 1mM Spermidine and 500 µg/ml BSA) and 2.5 units of T4-ligase (P.L. Biochem.). Ligation was carried out at 15°C for 12-18 hours.

b) "Block Ligation"

In the case of subassembly 2, the six internal oligonucleotides (i.e. all except the two that would contribute

the ultimate 5'-ends) were individually phosphorylated with [³²P]-ATP in a larger scale reaction, as described above. The six phosphorylated, labelled oligonucleotides were mixed, in equal proportions, with the remaining two, non-phosphorylated oligonucleotides, and were annealed. Ligation was carried out in the usual fashion, except that (1) the reaction mixtures contained at least a 5-fold higher concentration of oligonucleotides (i.e. 20 pmoles each in a final volume of 40 µl), and (2) no vector was added. This modified ligation mixture was electrophoresed through an 8% polyacrylamide gel. The correctly ligated "block" was identified on an autoradiogram by virtue of its size, compared to standard molecular weight markers. The gel band was excised, the DNA within it was electroeluted, and the material was concentrated and purified. The purified "block" was then mixed with an equimolar amount of the appropriate vector, and ligation was carried out in ligation buffer as described above.

Transformation

Frozen, competent cells of E. coli strain JM109 were prepared, and transfected by standard procedures (e.g. Maniatis et al., 1982).

Analysis of Transformants

Putative transformants were analyzed by performing a variation of the "mini-screen" described by Holmes and Quigley (1981). Those candidates which looked correct were further analyzed by full sequence analysis by the dideoxy chain termination method of Sanger et al. (1977).

Assembly of the 5' Half of the Prochymosin Gene

Subassembly pCHYM 1BB, illustrated in Figure 3, was formed from oligonucleotides 1B and 1B (the sequences of these

and all other component oligonucleotides are shown in Figure 2). The annealed oligonucleotides formed a Pst I site on one end, and an Eco RI site on the other, and were cloned into M13mpl9 that had been digested with both these enzymes. The prochymosin gene portion is located between the Pst I site and an Xba I site which lies just ahead of the Eco RI cloning site, so that the fragment could eventually be isolated by digestion with Pst I and Xba I. (Figure 3)

Subassembly pCHYM 2-4 (Figure 3) was formed from oligonucleotides 2A, 2B, 2C, 2D, 2 α , 2 β , 2 γ , and 2 δ . These were annealed, ligated and purified from an acrylamide gel in a "block" as described above. The completed block is bounded by the restriction sites for Xba I and Eco RI, and was cloned into M13mpl9 digested with these enzymes. The relevant portion of the prochymosin gene lies between the Xba I site and an Xma I site just ahead of the Eco RI cloning site.

The actual 5'-end of the prochymosin gene was provided in two alternate sets of oligonucleotides, 1A and 1 α as well as 1A(Y) and 1 α (Y), each pair of which has one end compatible with Bgl II and one end compatible with Pst I (Figure 3). The only difference is that the oligonucleotides with the "Y" designation have two additional nucleotides which have been added to ensure joining in the correct reading frame when using the yeast expression and secretion vector, pMV-1, as well as with the Aspergillus expression and secretion vector pGL2C. The alternative 5' end, comprised of oligos 1A and 1 α will ensure the correct reading frame when using the Aspergillus expression and secretion vector pAlcAlS.

To form the larger assembly pCHYM 1A(Y)B2 or pCHYM 1AB2-3, the annealed oligonucleotide pair representing the 5' end (either 1A(Y)+1 α (Y) or 1A+1 α), the Xba I/Pst I fragment from pCHYM 1B3, the Xba I/Eco RI fragment from pCHYM2-4, and M13mpl10 which had been digested with Bam HI and Eco RI, were

mixed in a molar ratio of 20:1:1:1 respectively, and were ligated in a single reaction.

The final, relevant prochymosin gene portion could therefore now be excised from this vector by digestion with Sau 3A and Xma I. (Figures 3 and 5)

Assembly of the 3'-Half of the Prochymosin Gene

Subassembly pCHYM 3-3 (Figure 4) was formed from the component oligonucleotides 3A, 3B, 3C, 3 α , 3 β , and 3 γ . The boundaries are delineated by an Xma I site and an Eco RI site. Cloning was done into M13mpl9 which had been similarly digested. The relevant prochymosin gene portion lies between the Xma I site and a Sal I site which lies just ahead of the Eco RI cloning site. (Figure 4)

Subassembly pCHYM 4-19 (Figure 4) contains the oligonucleotides 4A, 4B, 4 α , and 4 β , cloned between the Eco RI and Hind III sites of M13mpl9. All of the inserted material is authentic prochymosin coding sequence. (Figure 4)

The final subassembly, pCHYM 5-9 was formed from the remaining component oligonucleotides. These oligonucleotides form a prochymosin gene fragment bounded by Eco RI and Bam HI sites. They were assembled into M13mpl9 which had been digested with these two enzymes. (Figure 4)

The component fragments from these 3 subassemblies were ligated into the larger structure, designated pCHYM 345-21. The Bam HI/Sal I fragment from pCHYM 3-3, the Sal I/Eco RI fragment from pCHYM 4-19 and the Eco RI/Hind III fragment from pCHYM 5-9 were all identified and isolated by electrophoresis through gels of low melting agarose. They were then ligated, in equimolar concentrations, in a single reaction, to pBR 322 which had been digested with Bam HI and Hind III.

In this way, the relevant prochymosin gene portion could be excised with a combination of Xma I and Bam HI. (Figures 4 and 5)

Expression Vectors from Yeast and Filamentous Fungi

The recipient secretion and expression vectors used in this work have already been fully described, and are shown in schematic form in Figure 5. The pMV-1 plasmid expression vector for Saccharomyces cerevisiae is based upon the melibiase promoter and secretion signal contained on plasmid p4 ATCC 53360. The filamentous fungal expression vectors are based upon either the Aspergillus niger glucoamylase promoter and secretion signal (plasmid vector pGL2C ATCC 53367), or the Aspergillus nidulans alcohol dehydrogenase promoter and a synthetic signal (plasmid vector pAlcA1s ATCC 53368).

Each of these plasmid vectors has either a Bgl II or Bam HI cloning site at the end of the DNA sequence which encodes their respective secretion signals. The 5'-termini of the prochymosin gene segments were designed to be compatible with either of these sites. (Figure 5)

Assembly of Prochymosin Expression Vector pMV-1/CHYM 105

Plasmid pCHYM 1A(Y)B2 was digested with Sau3A and Xma I, while plasmid pCHYM 345-21 was digested with Xma I and Bam HI. In both cases the prochymosin-containing fragments were isolated and purified by electrophoresis through a gel of low-melting temperature agarose (2%). The yeast secretion/expression vector pMV-1 was digested with Bgl II, the 5'-phosphates were removed by treatment with bacterial alkaline phosphatase (IBI). The linear, dephosphorylated form was purified by electrophoresis through a gel of low-melting temperature agarose (1%). Aliquots of the three melted gel fragments were mixed to give approximately equimolar proportion

of fragments, the ligation was carried out at 15° C for 18 hrs. The resultant prochymosin expression/secretion vector, pMV-1/CHYM 105 ATCC 67294 (Figure 6), was identified by restriction analysis, and confirmed by full dideoxy nucleotide sequence analysis.

Assembly of Prochymosin Expression Vector pGL2c/CHYM 101

Plasmid pCHYM 1A(Y)B2 was digested with Sau3A and Xma I, while plasmid pCHYM 345-21 was digested with Xma I and Bam HI. In both cases the prochymosin-containing fragments were isolated and purified by electrophoresis through a gel of low-melting temperature agarose (2%). The Aspergillus niger secretion/expression vector pGL2C ATCC 53367 was digested with Bgl II, the 5'-phosphates were removed by treatment with bacterial alkaline phosphatase (IBI). The linear, dephosphorylated form was purified by electrophoresis through a gel of low-melting temperature agarose (1%). Aliquots of the three melted gel fragments were mixed to give approximately equimolar proportion of fragments, and ligation was carried out at 15° C for 18 hrs. The resultant prochymosin expression/secretion vector, pGL2C/CHYM 101 ATCC 67296 (Figure 7), was identified and confirmed by restriction analysis.

Assembly of Prochymosin Expression Vector pAlcAlSL/CHYM 103

Plasmid pCHYM 1A(Y)B2 was digested with Sau3A and Xma I, while plasmid pCHYM 345-21 was digested with Xma I and Bam HI. In both cases the prochymosin-containing fragments were isolated and purified by electrophoresis through a gel of low-melting temperature agarose (2%). The Aspergillus nidulans secretion/expression vector pAlcAlS ATCC 53368 was digested with Bam HI, the 5'-phosphates were removed by treatment with bacterial alkaline phosphatase (IBI). The linear, dephosphorylated form was purified by electrophoresis through a gel of low-melting temperature agarose (1%). Aliquots of the

three melted gel fragments were mixed to give approximately equimolar proportion of fragments, and ligation was carried out at 15° C for 18 hrs. The resultant prochymosin expression/secretion vector, pAlcAlsL/CHYM 103 ATCC 67295 (Figure 8), was identified and confirmed by restriction analysis.

Transformation into Yeast and Filamentous Fungi

Transformation of a Ura⁻ yeast strain SC295 to Ura⁺ using expression vector pMV-1/CHYM 105 was carried out as previously described (Ito et al., 1984).

Co-transformation of ArgB- Aspergillus niger or Aspergillus nidulans with an ArgB⁺ marker plus either of pGL2c/CHYM 101 or pAlcAlsL/CHYM 103 was performed as previously described (Buxton et al., 1985; Ballance et al., 1983). Positive transformants were identified by their ability to hybridize to nick-translated Bgl II/Eco RV fragment from plasmid pGL2c/CHYM 103, in a modified colony hybridization.

Expression of the Synthetic Prochymosin Gene by Yeast

a) Expression of Prochymosin-Specific Messenger RNA

Polyadenylated mRNA was isolated, by standard practices, both from yeast cells that had been transformed with the prochymosin expression plasmid, pMV-1/CHYM 105, and from the untransformed parental cells. The mRNA was blotted onto GeneScreen membranes (NEN) according to manufacturer's instructions. The blots were hybridized to nick-translated probes derived from plasmids containing portions of the synthetic prochymosin gene. Strong positive signals were detected from the transformed strains, when grown in the absence of glucose. Considerably weaker signals were detected in the transformed strains that had been grown in medium containing glucose, a condition which is well documented as being

repressing for the melibiase promoter which is herein used to control the expression of the prochymosin gene (e.g. Friis and Ottolenghi, 1959). By comparison, no hybridization was detected in the untransformed parent, in either growth condition. This indicates that the transformed strain is indeed expressing a messenger RNA which is specific for the synthetic prochymosin coding region, and that the expression of the synthetic gene is capable of being controlled by appropriate biological switches.

b) Immunological Detection of Prochymosin Antigens

Polyclonal antiserum directed against commercially prepared authentic bovine chymosin was derived, from rabbits, by standard procedures. The anti-chymosin antiserum was used in an ELISA test, utilizing urease-conjugated anti-rabbit IgG as a test system (Allelix UREIASE Reagents). Untransformed parents and transformed cells were grown to stationary phase, and the medium was concentrated by ultrafiltration through an Amicon P-10 membrane. Samples of the concentrated media were tested by the ELISA for the presence of (pro)chymosin-specific antigen. Once again, the transformed cells produced a positive response whereas the untransformed parents showed only a weak (background) response. This indicates that prochymosin is indeed being secreted from the yeast cells and can be found in the extracellular medium.

c) Biological Milk-Clotting Activity

As a final test of prochymosin activity, the concentrated, extracellular medium from cells grown to stationary phase was tested for the ability to clot milk. The prochymosin secreted from these cells must first be processed to an active form by incubation at acid pH (e.g. Pederson et al., 1979; Foltmann, 1979). Incubation at pH 2 produces pseudochymosin, whereas incubation at pH around 4-4.5 produces chymosin. Therefore, the extracellular supernatants were

activated by titration to the desired pH, followed by neutralization. Alternatively, cells were grown in medium at a pH already low enough (e.g. pH 3.5) to cause activation of any secreted prochymosin. Clotting was determined by a method similar to that of Emtage et al., 1983, using skim milk powder (Difco) reconstituted in 10 mM CaCl_2 (10% w/v). Using this assay, concentrated medium from transformed cells was shown, after activation, to contain an enzyme activity which was able to clot milk. The medium from the untransformed parent showed no similar activity.

Expression of the Synthetic Prochymosin Gene by Aspergillus

a) Expression of Prochymosin-Specific Messenger RNA

In an experiment similar to that described above for yeast cells, Aspergillus nidulans cells that had been transformed with the prochymosin expression vector pGL2c/CHYM 101 were tested for the presence of mRNA which would hybridize to a prochymosin-specific probe. A nick-translated DNA fragment from the synthetic prochymosin gene was shown to hybridize, with varying degrees of strength, to polyadenylated mRNA isolated from transformed cells. The differences in strength of hybridization probably reflect differences in copy number of the integrated genes. By comparison, no hybridization was observed with mRNA from untransformed parental cells.

b) Immunological Detection of Prochymosin Antigens

In an experiment similar to that described above for yeast cells, Aspergillus nidulans cells that had been transformed with the prochymosin expression vector pAlcAlSL/CHYM 103 were tested for the presence of prochymosin in the (concentrated) extracellular medium, by using the same anti-chymosin antibody and UREIASE-based ELISA methods. Once again, there was a positive reaction when using medium in which

transformed cells had been grown, but not from medium in which the untransformed cells had been grown.

While specific reference is made herein to an optimized prochymosin coding region, those skilled in the art will readily appreciate that the same principle, of optimizing codon usage in a coding region in accordance with the codon bias of a particular recipient cell, is applicable to protein coding regions other than prochymosin.

In addition, it will be appreciated that although the coding region may be synthesized specifically with a view to satisfying the codon preference in a given host, the coding region can be expressed by hosts other than the given host. As exemplified herein, the yeast optimized prochymosin coding region is also expressed by filamentous fungi. Accordingly, the utility of the optimized coding regions is not necessarily limited to transforming one particular host species or genus.

It will be further appreciated that the vectors specifically disclosed herein are described for the purpose of exemplification. Further modification to the vector components such as the promoter region, the signal sequence and other components, may be carried out in order to enhance the level at which prochymosin is expressed and/or secreted by a transformed host. Once produced by the host, prochymosin can be converted readily to chymosin using standard procedures.

Alternatively, vectors may be prepared as described which incorporate the optimized chymosin region only (as opposed to the entire region which encodes prochymosin). In that case, the "pro-" region may be deleted or may be modified in such a manner as to enhance expression of the chymosin enzyme by the transformed host.

REFERENCES CITED

- ABI User Bulletin No. 13 (1984) "Evaluation and Purification of Synthetic Oligonucleotides", Applied Biosystems Inc., Foster City, Calif.
- Ballance, D.J. et al (1983) Biochem. Biophys. Res. Commun. 112, 284-289.
- Beaucage, S.L. and M.H. Carruthers (1981) Tetrahedron Letters 22, 1859-1862.
- Bennetzen J.L. and B.D. Hall (1982) J. Biol. Chem. 257, 3026-3031.
- Burke, R.L. et al. (1983) J. Biol. Chem. 258, 2193-2201.
- Buxton, F.P. et al. (1985) Gene 37, 207-214.
- Devereux et al. (1984) Nucl. Acids Res. 12, 387-395.
- Emtage, J.S. et al. (1983) Proc. Nat. Acad. Sci (USA) 80, 3671-3675.
- Foltmann, B. (1979) in "Methods in Enzymology", Vol. 19 (Perlman G.E. and Lorand, L., eds), Academic Press, N.Y., pp421-436.
- Foltmann, B. et al. (1979) J. Biol. Chem. 254, 8447-8456.
- Friis, J. and P. Ottohenghi (1959) Compte Rendue Lab. Carlsberg 31, 272.
- Goff, C.G. et al. (1984) Gene 27, 35-46.
- Gouy, M. and C. Gautier (1982) Nucl. Acids Res. 10, 7055-7074.
- Harris, T.J.R. et al. (1982) Nucl. Acids Res. 10, 2177.
- Hitzeman, R.A. et al, (1982) Nucl. Acids Res. 10, 7791-7808.
- Holland, M.J. et al., (1981) J. Biol. Chem. 256, 1385-1395.
- Holland, J.P. and Holland, M.J. (1987) J. Biol. Chem. 254, 9839-9845.
- Holmes, D.S. and M. Quigley (1981) Analytical Biochem. 114, 193-197.
- Ikemura, T. (1982) J. Mol. Biol. 82, 573-597.
- Ito, H. et al (1984) Agric. Biol. Chem. 48, 341.
- Maniatis, T. et al., (1982) "Molecular Cloning - A Laboratory Manual", Cold Spring Harbor, N.Y.

- Marston, F.A.O. et al. (1984) Bio/Technology 2, 800-804.
- McBride, L.J. and M.H. Carruthers (1983) Tetrahedron Letters 24, 245-248.
- Moir, D. et al. (1982) Gene 19, 127-138.
- Moir, D. et al. (1985) in "Developments in Industrial Microbiology", Society for Industrial Microbiology, Proceedings, Chapter 2.
- Nishimori, K. et al. (1982) J. Biochem 91, 1085-1088.
- Nishimori, K. et al. (1984) Gene 29, 41-49.
- Pedersen, V.B. et al. (1975) Eur. J. Biochem. 55, 95-103.
- Pedersen, V.B. et al. (1979) Eur. J. Biochem. 94, 573-580.
- Sanger, F. et al. (1977) Proc. Nat. Acad. Sci. (USA) 74, 5463.
- Smith, R.A. et al. (1985) Science 229, 1219-1224.

CLAIMS:

1. A chymosin-encoding DNA segment having a codon sequence optimized for expression in a foreign host.
2. A prochymosin-encoding DNA segment having a codon sequence optimized for expression in a foreign host.
3. The DNA segment according to claim 1 or claim 2 wherein said host is selected from yeast and filamentous fungi.
4. The DNA segment according to claim 3 wherein said host is Saccharomyces sp.
5. The DNA segment according to claim 3 wherein said host is Aspergillus sp.
6. An expression vector comprising the DNA segment as defined in claim 1 or claim 2.
7. The expression vector according to claim 6 which comprises a promoter region operatively coupled with said DNA segment.
8. The expression vector according to claim 7 which comprises a signal sequence operatively associated with said DNA segment.
9. The expression vector according to claim 8 which comprises the promoter region and signal sequence of the melibiase gene of Saccharomyces sp.
10. The expression vector according to claim 8 which comprises the promoter region and signal sequence of the glucoamylase gene of Aspergillus sp.

11. The expression vector according to claim 8 which comprises the promoter region and signal sequence defined on plasmid pAlcA1S ATCC 53368.
12. The expression vector according to claim 8 which is selected from the plasmids pGL2C/CHYM 101, ATCC 67296; pAlcA1S/CHYM 103, ATCC 67295 and pMV-1/CHYM 105, ATCC 67294.
13. Transformed cells comprising a DNA segment defined in claim 1 or claim 2.
14. Transformed cells according to claim 13 which are yeasts.
15. Transformed cells according to claim 14 which are of the species S. cerevisiae.
16. Transformed cells according to claim 13 which are filamentous fungi.
17. Transformed cells according to claim 16 which are of the genus Aspergillus.
18. Transformed cells according to claim 17 which are of the species Aspergillus nidulans or Aspergillus niger.
19. A process for producing prochymosin which comprises culturing transformed cells as defined in claim 13 under growth promoting conditions.
20. A process for producing chymosin which comprises culturing transformed cells as defined in claim 13 under growth promoting conditions, and converting the prochymosin expressed thereby to chymosin.

